

(1)

The blastp program is used because we are comparing protein sequences from the human.fa file against the mouse protein database. Since both input and target sequences are proteins, blastp is the appropriate tool for this type of sequence comparison.

(2)

A substitution matrix, such as BLOSUM62, is used to score alignments between protein sequences based on the likelihood of amino acid substitutions occurring. BLOSUM62 is commonly used for general-purpose protein sequence comparison because it provides a balance between sensitivity and specificity for alignments of moderately diverged proteins.

(3)

The E-value threshold is set to 0.001 to ensure only statistically significant alignments are considered, reducing the number of false positives by focusing on matches that are less likely to occur by chance.

The output format is set to XML (outfmt=5), which is easy to parse programmatically using BioPython's NCBI XML parser and provides detailed information about the alignments.